Verdichtete Informationen

Maßzahlen

Statistiken
bei Stichproben

Parameter
bei Grundgesamtheiten

Maßzahlen
zur Beschreibung
univariater Verteilungen

Maßzahlen der
zentralen Tendenz
(Mittelwerte)

Maßzahlen der
Variabilität
(Streuungswerte)

MAßZAHLEN DER ZENTRALEN TENDENZ

Eine Maßzahl der zentralen Tendenz (im Benninghaus: Mittelwert) ist der Kennwert, der die gesamte Verteilung am besten repräsentiert.

1. Modus (h)

Der Modus ist der Wert, der in einer Verteilung am häufigsten vorkommt (dichtester Wert).

Beispiel:

5, 6, 6, 7, 7, 7, 8, 8, 9, 10

h = 7

Beispiel:

5, 6, 6, 7, 7, 7, 8, 8, 8, 9, 10

h = 7,5 aufgrund von benachbarten Häufigkeitsmaxima

Treten zwei nicht benachbarte Messwerte mit relativen Häufigkeitsmaxima auf, dann werden diese beiden Messwerte als Modalwerte angegeben. Die Verteilung ist dann bimodal.

Beispiel:

x_i	f_i
4	1
5	2
6	4
7	2
8	1
9	5
10	1
11	1

h = 6, 9

2. Median (Zentralwert) $\tilde{x}$

Der Median ist der Wert, der eine nach ihrer Größe geordnete Reihe von Messwerten halbiert.

Die Bestimmung des Median ist abhängig von der Anzahl der Fälle:

a) liegt eine <u>ungerade</u> Anzahl von Fällen vor, so ist der Median der Wert des mittleren Falles:

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)}$$

<u>Beispiel:</u> gegeben sei eine Stichprobe von monatlichen Einkommen (in Euro) von Studenten

637	697	719	$\boxed{750}$	898	912	981

Da die Stichprobengröße n = 7 ist, ist der Median

$$\tilde{x} = x_{\left(\frac{7+1}{2}\right)} = x_4 = 750$$

b) liegt eine <u>gerade</u> Anzahl von Fällen vor, so ist der Median der Wert, der die beiden mittleren Fälle halbiert:

$$\tilde{x} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

<u>Beispiel:</u> gegeben sei eine Stichprobe von monatlichen Einkommen (in Euro) von Studenten

$$637 \quad 697 \quad 719 \quad \underbrace{750 \quad 898} \quad 912 \quad 981 \quad 1032$$

$$\tilde{x} = \frac{x_{\left(\frac{8}{2}\right)} + x_{\left(\frac{8}{2}+1\right)}}{2} = \frac{x_4 + x_5}{2} = \frac{750 + 898}{2} = 824$$

Median bei klassifizierten Häufigkeiten:

1) Man bildet die kumulierten Häufigkeiten.

2) Man ermittelt n/2 bzw. n+1/2. Damit ist die Stellung des Medians festgelegt.

3) Man bestimmt aufgrund der kumulierten Häufigkeit diejenige Messwertklasse, in die der Median fällt (Eingriffsspielraum).

4)Man formuliert die exakten Grenzen des Eingriffsspielraumes.

5) Man berechnet den Median nach der Formel:

$$\tilde{x} = U + \frac{\left(\frac{N}{2}\right) + FU}{Fm} \times h$$

wobei: U = exakte untere Grenze des Eingriffsspielraumes (Medianintervall)
N = Anzahl der Fälle
FU = kumulierte Häufigkeit unterhalb des Medianintervalls
Fm = Häufigkeit im Medianintervall
h = Klassenbreite

Beispiel:

1) Man bildet die kumulierte Häufigkeit (cum f_i)

4) exakte Grenze 11,5

Klassenintervall	f_i	cum f_i
6 – 8	3	3
9 – 11	10	$\boxed{13}$
12 – 14	14	27
15 – 17	13	40
18 – 20	11	51

3) Eingriffs-spielraum (Median-intervall)

2) Man ermittelt die Stellung des Medians:

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{51+1}{2}\right)} = x_{(26)}$$ der 26. Wert

5) $\tilde{x}$ = Exakte untere Grenze des Eingriffsspielraumes (Medianintervalls) + $\left(\dfrac{\text{N/2 – cum } f_i \text{ unterhalb des Medianintervalls}}{f_i \text{ im Medianintervall}} \right)$ * h

$$\tilde{x} = 11{,}5 + \left(\frac{51/2 - 13}{14} \right) * 3$$

$$\tilde{x} = 14{,}18$$

3. Das arithmetische Mittel ($\bar{x}$)

Das arithmetische Mittel (Durchschnittswert) ist definiert als die Summe der Messwerte, geteilt durch ihre Anzahl:

$$\bar{x} = \frac{x_1 + x_2 + x_3 ... + x_n}{n} = \frac{\sum_{i=1}^{N} x_i}{n}$$

Liegt mehr als eine Häufigkeit vor:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 ... + f_k x_k}{n} = \frac{\sum_{i=1}^{K} f_i x_{i_i}}{n}$$

Beispiel: Sanierungsgebiet

x_i = Alter der Häuser in Jahren

x_i	f_i
30	1
40	9
50	7
100	2
400	1
Σ	20

$$\bar{x} = \frac{(1 \times 30) + (9 \times 40) + (7 \times 50) + (2 \times 100) + (1 \times 400)}{20}$$

$$\bar{x} = \frac{(30) + (360) + (350) + (200) + (400)}{20}$$

$$\bar{x} = \frac{1340}{20} = 67$$

ohne Ausreißer:

$$\bar{x} = \frac{(1 \times 30) + (9 \times 40) + (7 \times 50) + (2 \times 100)}{19} = \frac{940}{19} = 49{,}47$$

Die Eigenschaften des arithmetischen Mittels:

1) Die Summe der Abweichungen aller Messwerte von ihrem arithmetischen Mittel ist gleich Null.

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right) \quad = 0$$

Beispiel: 5, 6, 7, 8, 9

$$\overline{x} = 7$$

x_i	$x_i - \overline{x}$
5	-2
6	-1
7	0
8	1
9	2
Σ	0

2) Die Summe der Quadrate der Abweichungen aller Messwerte von ihrem arithmetischen Mittel ist kleiner als die Summe der Quadrate der Abweichungen aller Messwerte von einem beliebigen Wert der Verteilung.

Mit anderen Worten: Die Summe der Abweichungsquadrate ist für das arithmetische Mittel ein Minimum.

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 \quad < \quad \sum_{i=1}^{n} (x_i - x_0)^2$$

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 \quad = \min!$$

:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad > \quad \sum_{i=1}^{n}(x_i-x_0)^2$$

Beispiel: 5,6,7,8,9

$$\bar{x}=7$$

angenommen: $x_0 = 6$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	x_i-x_0	$(x_i-x_0)^2$
5	-2	4	-1	1
6	-1	1	0	0
7	0	0	1	1
8	1	1	2	4
9	2	4	3	9
$\sum$	0	10	5	15

Also: 10 < 15

3) Die Addition (Subtraktion) einer bestimmten Zahl zu allen Einzelwerten einer Verteilung vergrößert (verkleinert) das arithmetische Mittel um diese Zahl.

$(x_i + c) \Rightarrow \overline{x} + c$

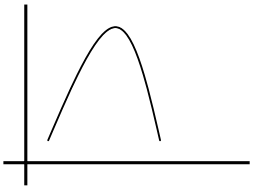
bzw. $(x_i - c) \Rightarrow \overline{x} - c$

Aufgaben:

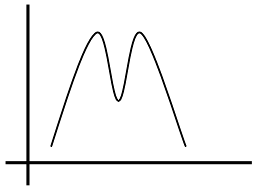
1. Elf Studenten nehmen an einem Elfmeterschießen teil. Sie erzielen folgende Trefferzahlen bei 10 Schüssen: 4, 6, 3, 1, 2, 8, 4, 5, 2, 0, 2. Bestimmen Sie Modalwert, Median, arithmetisches Mittel.
2. In einem Flugblatt wird verkündet: Bei zwei Umfragen unter Sudenten haben sich einmal 60% von 100 Hörern einer Vorlesung und zum anderen 38% von 1000 vor der Mensa befragten Studenten für die Abschaffung der Statistik ausgesprochen. Wie viel Prozent der Befragten haben sich im Durchschnitt für die Abschaffung der Statistik ausgesprochen?

3) Welche Maßzahl der zentralen Tendenz ist angemessen?

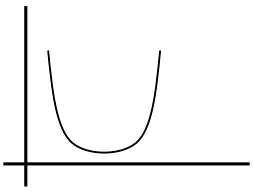
1)



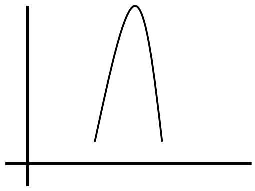
2)



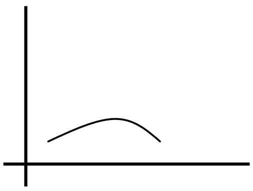
3)



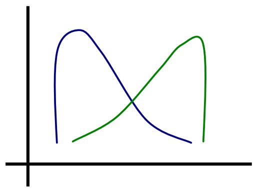
4)



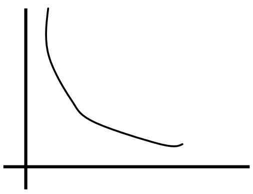
5)



6)



7)



Maßzahlen der zentralen Tendenz und Skalenniveau

Maßzahlen der zentralen Tendenz				
		Modus h	Median $\tilde{x}$	arithm. Mittel $\bar{x}$
Skalen-niveau	Nominal	X		
	Ordinal	X	X	
	Intervall	X	X	X

Merke:

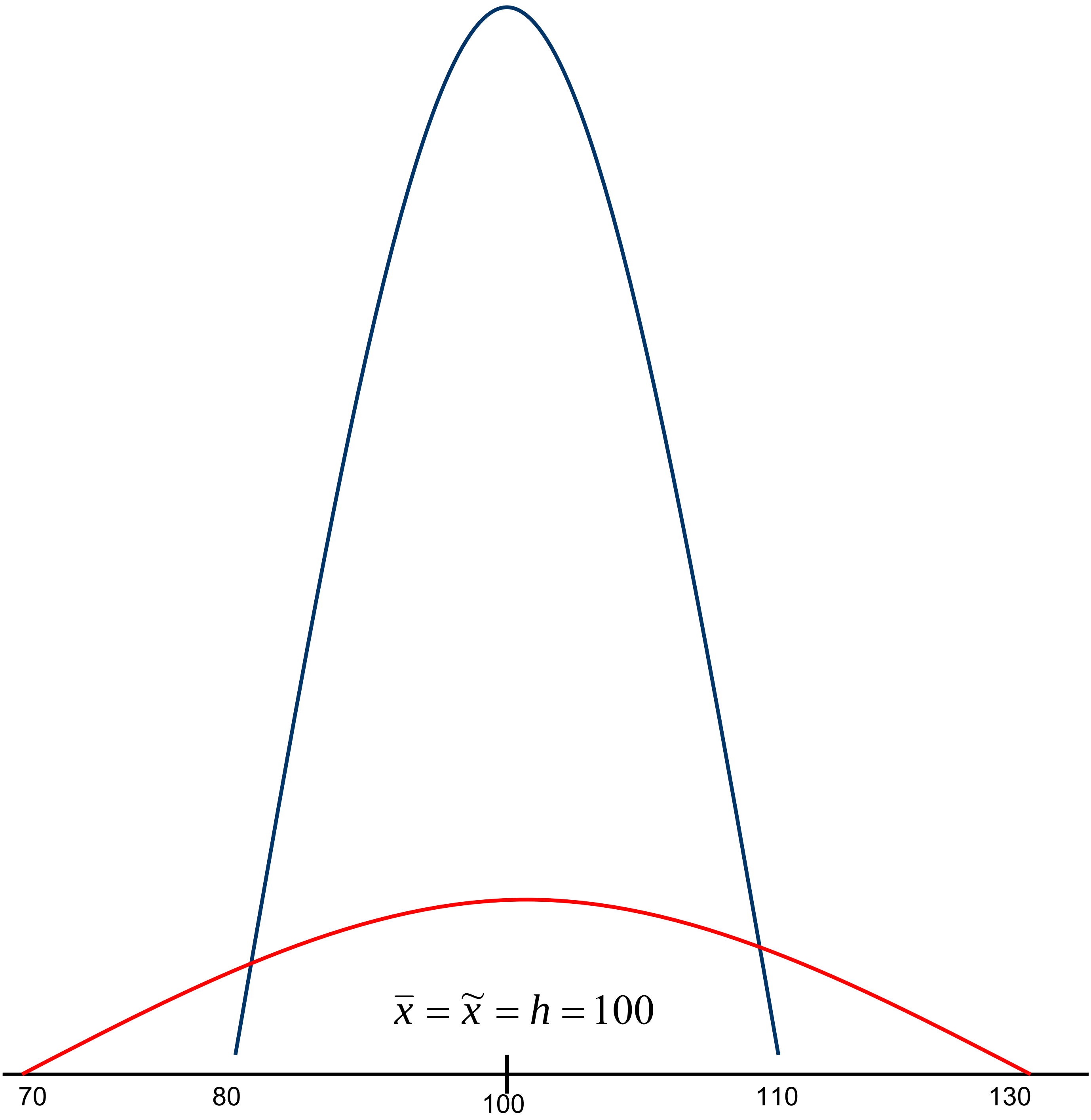
Maßzahlen der zentralen Tendenz

- Modus (h): typischer Wert einer Verteilung
- Median ($\tilde{x}$): zentraler Wert einer Verteilung
- Arithmetisches Mittel ($\bar{x}$): Durchschnittswert

Aufgabe: Bestimmen Sie den Modus, Median und arithmetisches Mittel!

x_i	f_i
1	2
3	5
7	3
11	2

Abbildung: Zwei Verteilungen mit gleicher zentraler Tendenz ($\bar{x} = \tilde{x} = h = 100$), aber ungleicher Streuung



STREUUNGSWERTE (VARIABILITÄTSMAßE)

Die Streuungswerte haben die Aufgabe, die Variabilität des gemessenen Merkmals zu beschreiben.

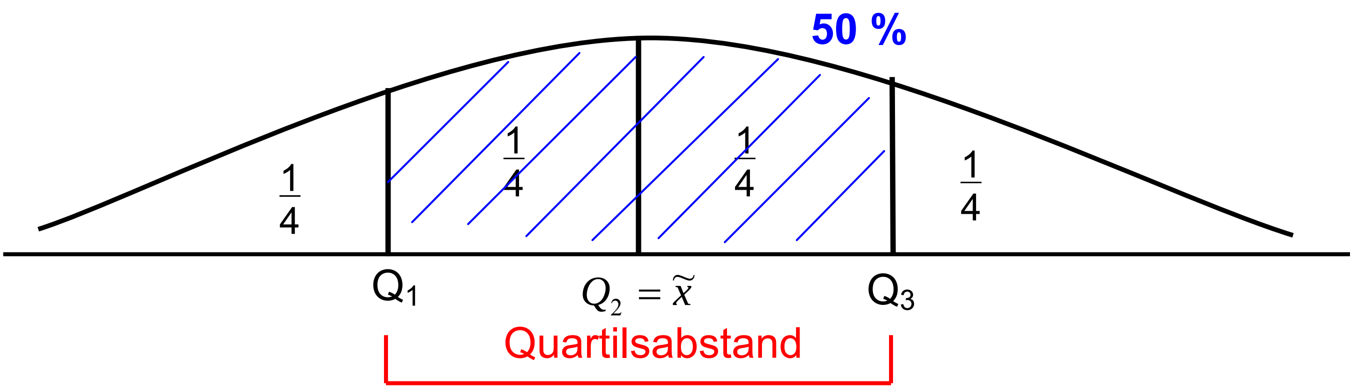
1. Der Range

Der Range ist definiert als die Differenz zwischen dem größten und dem kleinsten Messwert einer Verteilung.

$R = x_{max} – x_{min}$

2. Der Quartilsabstand und der mittlere Quartilsabstand

Abbildung: Illustration der Quartile und des Quartilsabstandes



Quartilsabstand = Q_3 - Q_1

Beispiel: Häufigkeit eines Schwips

		x_i	f_i	cum f_i	
Nie		1	14	$\boxed{14}$	
Ganz selten	1,5	2	84	98	¼ N = 65 1. Quartilintervall
Manchmal		3	60	$\boxed{158}$	
Recht häufig	3,5	4	59	217	¾ N = 195 2. Quartilintervall
Regelmäßig		5	43	260	

1) Man ermittelt ¼ N bzw. ¾ N

2) Man bestimmt anhand der kumulierten Häufigkeitsverteilung die Quartilintervalle

3) Man vergewissert sich der exakten unteren Grenzen der Quartilintervalle (1,5 bzw. 3,5)

$$Q_1 = U + \frac{\frac{1}{4} N - FU}{Fm} * h \quad = \quad 1{,}5 + \frac{65 - 14}{84} * 1 \quad = 2{,}11$$

$$Q_3 = U + \frac{\frac{3}{4} N - FU}{Fm} * h \quad = \quad 3{,}5 + \frac{195 - 158}{59} * 1 \quad = 4{,}13$$

$$QA = Q_3 - Q_1 \quad = \quad 4{,}13 - 2{,}11 = 2{,}02$$

---

wobei:

U = exakte untere Grenze des Quartilsintervalls

FU = kumulierte Häufigkeit unterhalb des Quartilsintervalls

Fm = Häufigkeit im Quartilsintervall

Aufgabe:

Gegeben sind Messwertpaare von 10 Objekten:

x: 60 65 70 70 77 78 79 82 84 135

y: 80 84 75 90 70 83 95 81 54 88

a) bestimmen Sie für beide Reihen den Modus

Reihe 1

Reihe 2

b) berechnen Sie für die erste Reihe den Median und für die zweite Reihe den Range (Spannweite)

Reihe 1

Reihe 2

3. Die durchschnittliche Abweichung

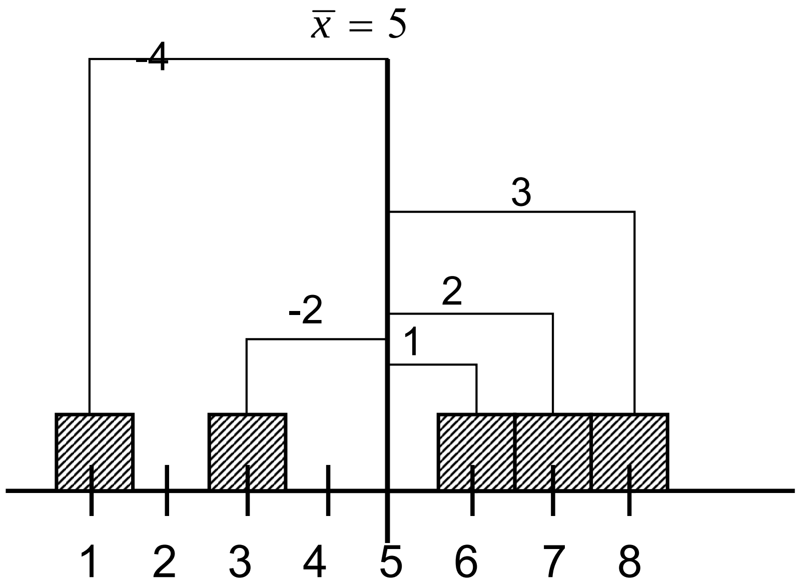
Die durchschnittliche Abweichung ist definiert als der Durchschnitt der absoluten Abweichungen aller Messwerte einer Verteilung von ihrem arithmetischen Mittel:

$$AD = \frac{\sum_{i=1}^{k} \left| x_i - \overline{x} \right|}{N}$$

bzw.:

$$AD = \frac{\sum_{i=1}^{k} \left| x_i - \overline{x} \right| \times f_i}{N}$$

Abbildung: Darstellung der Abweichungen der Messwerte von ihrem arithmetischen Mittel



$$AD = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right| \times f_i}{N}$$

$$\overline{x} = 62{,}93$$

x_i	f_i	$x_i - \overline{x}$	$\left| x_i - \overline{x} \right|$	$\left| x_i - \overline{x} \right| \times f_i$
48	8	-14,93	14,93	119,44
53	8	-9,93	9,93	78,44
58	16	-4,93	4,93	78,88
63	14	0,07	0,07	0,98
68	9	5,07	5,07	45,68
73	12	10,07	10,07	120,84
78	2	15,07	15,07	30,14
83	4	20,07	20,07	80,28
$\sum$	73			555,63

$$AD = \frac{555{,}63}{73} = 7{,}6$$

4. Varianz und Standardabweichung

a) Varianz

Die Varianz (s^2 bei Stichproben; σ^2 bei Grundgesamtheiten) ist die Summe der Abweichungsquadrate (quadrierte Abweichungen aller Messwerte einer Verteilung von ihrem arithmetischen Mittel, dividiert durch die Anzahl der Messungen.

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N} \qquad s^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2 \times f_i}{N}$$

bzw.

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1} \qquad s^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2 \times f_i}{N-1}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2 \times f_i}{N}$$

b) <u>Standardabweichung</u>

Die Standardabweichung(s bei Stichproben; σ bei Grundgesamtheiten) ist definiert als die Quadratwurzel aus der Varianz

$$s = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2}{N}} \qquad s = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2 \times f_i}{N}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2 \times f_i}{N}}$$

$$s = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2}{N-1}} \qquad s = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2 \times f_i}{N-1}}$$

Beispiel: Klassifizierte Gewichtsverteilung einer Studentengruppe im Kurs Statistik

x_i	f_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 \times f_i$
48	8	-14,93	222,900	1783,20
53	8	-9,93	98,600	788,80
58	16	-4,93	24,300	388,80
63	14	0,07	0,005	0,07
68	9	5,07	25,700	231,30
73	12	10,07	101,400	1216,80
78	2	15,07	227,100	454,20
83	4	20,07	402,800	1611,20
$\sum$	73			6474,37

Berechnung der Varianz:

$$s^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 \times f_i}{N}$$

$$s^2 = \frac{6474,37}{73} = 89,92$$

Berechnung der Varianz bei klassierten Daten

Beispiel:

Gegeben ist die folgende Häufigkeitsverteilung eines klassierten Merkmals:

Klasse	von 10 bis unter 20	von 20 bis unter 30	von 30 bis unter 40	von 40 bis unter 50
Häufigkeit	12	23	20	50

Klasse	**Klassen-mitte**	$\mathbf{f_i}$	$\mathbf{f_i x_i}$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 \times f_i$
10 – 19,9	15	12	180	-13	169	2028
20 – 29,9	25	23	575	-3	9	207
30 – 39,9	35	20	700	7	49	980
40 – 49,9	45	5	225	17	289	1445
Σ		**60**	**1680**			**4660**

$$\overline{x} = \frac{1680}{60} = 28 \qquad s^2 = \frac{4660}{60} = 77{,}67$$

$$s = \sqrt{77{,}67} = 8{,}81$$

5. Der Variationskoeffizient

$$v = \frac{s}{\overline{x}}$$

Zum Zweck des Vergleichs von Streuungen aus verschiedenen Grundgesamtheiten oder Stichproben benötigt man ein dimensionsloses Maß der Streuung. Eins solchen Maß ist der Variationskoeffizient.

Beispiel: Mittelwert und Standardabweichung einer CD in der BRD betragen

$$\overline{x} = 3{,}25 \text{ €}; \quad s = 1{,}4 \text{ €}$$

in Großbritannien $\overline{x} = 19\,\$$; $s = 8{,}18\,\$$

In beiden Ländern streut, gemessen am Durchschnitt, der Preis einer Cd ungefähr gleich, da gilt:

$$v = \frac{1{,}4}{3{,}25} \approx \frac{8{,}18}{19}$$

Aufgaben:

1. Für Waschpulver eines bestimmten Herstellers wurden in 10 Geschäften Essens folgende Preise für 1 kg Paket ermittelt (in Euro):
1,40; 1,60; 1,70; 1,50; 1,40; 1,80; 1,70; 1,60; 1,50; 1,80
Berechnen Sie Varianz und Standardabweichung!

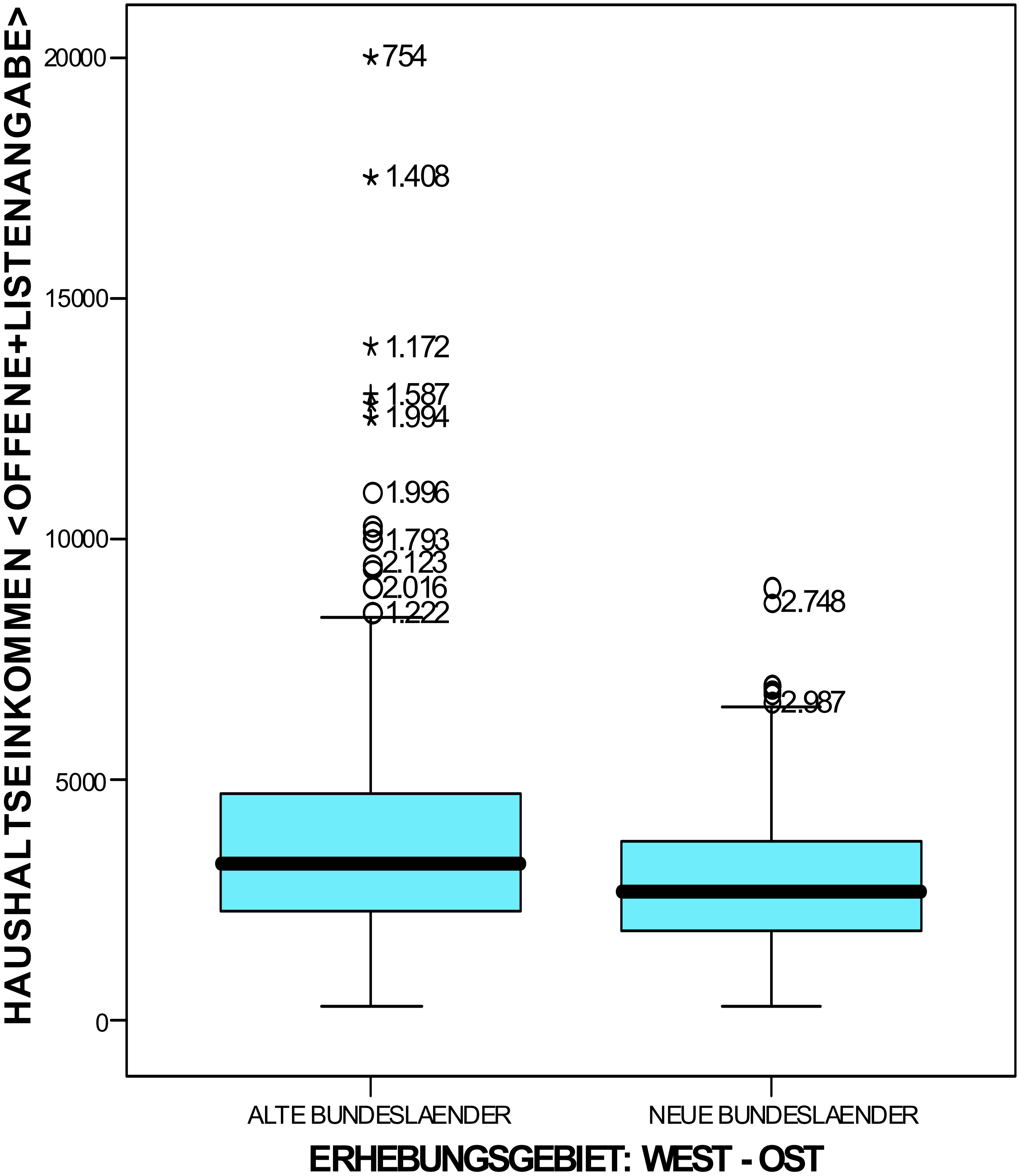
2. Die Untersuchung des verfügbaren Einkommens von Studenten in den USA und in der BRD ergab folgende Werte:

USA: $\overline{x} = 470\,\$$; $s = 160\,\$$

BRD: $\overline{x} = 720\ Euro$; $s = 180\ Euro$

Vergleichen Sie die relative Streuung miteinander!

Boxplot


HAUSHALTSEINKOMMEN <OFFENE+LISTENANGABE>
20000
15000
10000
5000
0
754
1.408
1.172
1.587
1.994
1.996
1.793
2.123
2.016
1.222
2.748
2.987
ALTE BUNDESLAENDER
NEUE BUNDESLAENDER
ERHEBUNGSGEBIET: WEST - OST